Randomized Matrix-Free Trace and Log-Determinant Estimators

Ilse Ipsen

Joint with: Alen Alexanderian & Arvind Saibaba

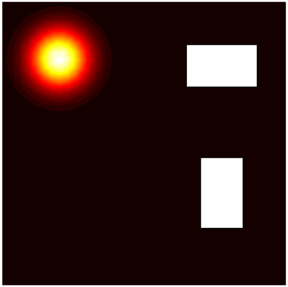
North Carolina State University
Raleigh, NC, USA

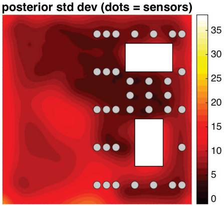
Research supported by DARPA XData

Our Contribution to this Minisymposium

Inverse problems: Bayesian OED

Inverse Problem: Diffusive Contaminant Transport





Forward problem: Time-dependent advection-diffusion equation
Inverse problem: Re-construct uncertain initial concentration
from measurements of 35 sensors, at 3 time points

Prior-preconditioned Fisher information $\mathcal{H} \equiv C_{\text{prior}}^{1/2} \, F^* \, \Gamma_{\text{noise}}^{-1} \, F \, C_{\text{prior}}^{1/2}$

Sensitivity of OED: $\operatorname{trace}(\mathcal{H})$

Bayesian D-optimal design criterion: $\log\det(\mathbf{I} + \mathcal{H})$

Our Contribution to this Minisymposium

Inverse problems: Bayesian OED
Big data: Randomized estimators

Given: Hermitian positive semi-definite matrix $\mathbf{A}$
Want: $\operatorname{trace}(\mathbf{A})$ and $\log\det(\mathbf{I}+\mathbf{A})$

Our Contribution to this Minisymposium

Inverse problems: Bayesian OED

Big data: Randomized estimators

Given: Hermitian positive semi-definite matrix $\mathbf{A}$

Want: $\text{trace}(\mathbf{A})$ and $\log\det(\mathbf{I} + \mathbf{A})$

Our estimators

- Fast and accurate for Bayesian inverse problem
- Matrix free, simple implementation
- Much higher accuracy than Monte Carlo
- Informative probabilistic bounds, even for small dimensions

Existing Estimators

Small Explicit Matrices: Direct Methods

Trace = sum of diagonal elements

$$\mathrm{trace}(\mathbf{A}) = \sum_j a_{jj}$$

Logdet

(1) Convert to trace

$$\log\det(\mathbf{I}+\mathbf{A}) = \mathrm{trace}(\log(\mathbf{I}+\mathbf{A}))$$

(2) Cholesky factorization $\mathbf{I}+\mathbf{A} = \mathbf{L}\mathbf{L}^*$

$$\log\det(\mathbf{I}+\mathbf{A}) = \log|\det(\mathbf{L})|^2 = 2\log\prod_j l_{jj}$$

Large Sparse or Implicit Matrices

Trace: Monte Carlo methods

$$\mathrm{trace}(\mathbf{A}) \approx \tfrac{1}{N} \sum_{j=1}^{N} \mathbf{z}_j^* \mathbf{A} \mathbf{z}_j$$

N independent random vectors $\mathbf{z}_j$
Rademacher, standard Gaussian $\Rightarrow$ unbiased estimator

Hutchinson 1989, Avron & Toledo 2011
Roosta-Khorasani & Ascher 2015, Lin 2016

Logdet: Expansion of log

$$\log\det(\mathbf{I} + \mathbf{A}) \approx \mathrm{trace}\left(\sum_j p_j(\mathbf{A}) \right)$$

Barry & Pace 1999, Pace & LeSage 2004, Zhang et al. 2008
Chen et al. 2011, Anitescu et al. 2012, Boutsidis et al. 2015
Han et al. 2015

Our Estimators

Randomized Subspace Iteration

Given: $n \times n$ matrix $\mathbf{A}$ with k dominant eigenvalues

Compute low rank approximation $\mathbf{T}$:

(1) Pick random starting guess $\mathbf{\Omega}$ with $\ell \geq k$ columns

(2) Subspace iteration $\mathbf{Y} = \mathbf{A}^q\mathbf{\Omega}$

(3) Orthonormalize: Thin QR $\mathbf{Y} = \mathbf{QR}$

(4) Compute $\ell \times \ell$ matrix $\mathbf{T} = \mathbf{Q}^*\mathbf{AQ}$

Estimators:

$$\mathrm{trace}(\mathbf{T}) \approx \mathrm{trace}(\mathbf{A}) \qquad \log\det(\mathbf{I} + \mathbf{T}) \approx \log\det(\mathbf{I} + \mathbf{A})$$

Analysis

Given: $n \times n$ hpsd matrix $\mathbf{A}$, $\ell \times \ell$ approximation $\mathbf{T}$

Want bounds for: $\operatorname{trace}(\mathbf{T})$ and $\log\det(\mathbf{I} + \mathbf{T})$

Ingredients:

- Eigenvalues of $\mathbf{A}$

$$\lambda_1 \geq \cdots \geq \underbrace{\lambda_k \quad \gg \quad \lambda_{k+1}}_{gap} \geq \cdots \geq \lambda_n \geq 0$$

- Number of subspace iterations q (typically $1-2$)
- Oversampling parameter $k \leq \ell \ll n$
- $n \times \ell$ random starting guess $\mathbf{\Omega}$

Analysis

Given: $n \times n$ hpsd matrix $\mathbf{A}$, $\ell \times \ell$ approximation $\mathbf{T}$

Want bounds for: $\operatorname{trace}(\mathbf{T})$ and $\log\det(\mathbf{I} + \mathbf{T})$

Ingredients:

- Eigenvalues of $\mathbf{A}$

$$\lambda_1 \geq \cdots \geq \underbrace{\lambda_k \gg \lambda_{k+1}}_{gap} \geq \cdots \geq \lambda_n \geq 0$$

- Number of subspace iterations q (typically $1-2$)
- Oversampling parameter $k \leq \ell \ll n$
- $n \times \ell$ random starting guess $\mathbf{\Omega}$

Derivation of bounds has 2 parts:

(1) Structural: Perturbation bound for any $\mathbf{\Omega}$

(2) Probabilistic: Exploit properties of random $\mathbf{\Omega}$

Structural Bounds for Trace

Deterministic bounds for any starting guess

Requirements: Gap and Subspace Contribution

Eigenvalue decomposition $\mathbf{A} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^*$

- Dominant eigenspace of dimension k

$$\boldsymbol{\Lambda} = \begin{pmatrix} \boldsymbol{\Lambda}_1 & \\ & \boldsymbol{\Lambda}_2 \end{pmatrix} \qquad \mathbf{U} = \begin{pmatrix} \mathbf{U}_1 & \mathbf{U}_2 \end{pmatrix}$$

Eigenvalues $\lambda_1 \geq \cdots \geq \lambda_k \gg \lambda_{k+1} \geq \cdots \geq \lambda_n \geq 0$

- Strong eigenvalue gap

$$\gamma \equiv \lambda_{k+1}/\lambda_k = \|\boldsymbol{\Lambda}_2\| \, \|\boldsymbol{\Lambda}_1^{-1}\| \ll 1 \qquad \text{(2-norm)}$$

- Starting guess close enough to dominant eigenspace

$$\mathbf{U}^*\boldsymbol{\Omega} = \begin{pmatrix} \mathbf{U}_1^*\boldsymbol{\Omega} \\ \mathbf{U}_2^*\boldsymbol{\Omega} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\Omega}_1 \\ \boldsymbol{\Omega}_2 \end{pmatrix} \qquad \operatorname{rank}(\boldsymbol{\Omega}_1) = k$$

Absolute Error

- Exact computation: If $\mathrm{rank}(\mathbf{A}) = k$ then

$$\mathrm{trace}(\mathbf{T}) = \mathrm{trace}(\mathbf{A})$$

- Perfect starting guess: If $\boldsymbol{\Omega} = \mathbf{U}_1$ then

$$0 \leq \mathrm{trace}(\mathbf{A}) - \mathrm{trace}(\mathbf{T}) = \underbrace{\mathrm{trace}(\boldsymbol{\Lambda}_2)}_{\textit{small eigenvalues}}$$

Absolute Error

- Exact computation: If $\text{rank}(\mathbf{A}) = k$ then

$$\text{trace}(\mathbf{T}) = \text{trace}(\mathbf{A})$$

- Perfect starting guess: If $\boldsymbol{\Omega} = \mathbf{U}_1$ then

$$0 \leq \text{trace}(\mathbf{A}) - \text{trace}(\mathbf{T}) = \underbrace{\text{trace}(\boldsymbol{\Lambda}_2)}_{\textit{small eigenvalues}}$$

- General starting guess:

$$0 \leq \text{trace}(\mathbf{A}) - \text{trace}(\mathbf{T}) \leq \left(1 + \gamma^{2q-1} \, \|\boldsymbol{\Omega}_2 \boldsymbol{\Omega}_1^{\dagger}\|^2\right) \text{trace}(\boldsymbol{\Lambda}_2)$$

Absolute Error

General starting guess

$$0 \leq \mathrm{trace}(\mathbf{A}) - \mathrm{trace}(\mathbf{T}) \leq \left(1 + \gamma^{2q-1} \, \|\mathbf{\Omega}_2 \mathbf{\Omega}_1^\dagger\|^2\right) \mathrm{trace}(\mathbf{\Lambda}_2)$$

Structural properties of estimator

- Limited by mass of subdominant eigenvalues, $\mathrm{trace}(\mathbf{\Lambda}_2)$
- Accurate if $\mathbf{A}$ has low numerical rank, $\mathbf{\Lambda}_2 \approx 0$
- Converges fast if
  Dominant eigenvalues well separated, $\gamma \ll 1$
  Starting guess close to dominant subspace, $\|\mathbf{\Omega}_2 \mathbf{\Omega}_1^\dagger\| \ll 1$

Probabilistic Bounds for Trace

Starting guess is Gaussian

Absolute Error: Expectation

If starting guess $\mathbf{\Omega}$ is $n \times (k+p)$ standard Gaussian

k : Dimension of dominant subspace

p : Oversampling parameter

then

$$0 \leq \mathbb{E}\left[\operatorname{trace}(\mathbf{A}) - \operatorname{trace}(\mathbf{T})\right] \leq \left(1 + c\,\gamma^{2q-1}\right) \operatorname{trace}(\mathbf{\Lambda}_2)$$

where

$$c = \underbrace{\frac{e^2}{p^2-1}\,\left(\frac{1}{4(p+1)}\right)^{\frac{2}{p+1}}}_{\leq .2} (k+p)\left(\sqrt{n-k} + 2\sqrt{k+p}\right)^2$$

Estimator is biased (for $\mathbf{\Lambda}_2 \neq 0$)

Absolute Error: Concentration

Starting guess $\mathbf{\Omega}$ is $n \times (k+p)$ standard Gaussian

For any $\delta > 0$, with probability at $1-\delta$

$$0 \leq \operatorname{trace}(\mathbf{A}) - \operatorname{trace}(\mathbf{T}) \leq \left(1 + c\,\gamma^{2q-1}\right) \operatorname{trace}(\mathbf{\Lambda}_2)$$

where

$$c \equiv \left(\frac{2}{\delta}\right)^{\frac{2}{p+1}} e^2 \, \frac{k+p}{(p+1)^2} \left(\sqrt{n-k} + \sqrt{k+p} + \sqrt{2\log\frac{2}{\delta}}\right)^2$$

Fast convergence with high probability:

Failure probability $\delta \approx 10^{-15}$
Iterations $q \leq 2$, oversampling $p \approx 30$
Strong eigenvalue gap $\gamma \leq .8$

Logdet Estimator

Absolute Error: Structural Bounds

- Exact computation: If $\mathrm{rank}(\mathbf{A}) = k$ then

$$\log\det(\mathbf{I}+\mathbf{T}) = \log\det(\mathbf{I}+\mathbf{A})$$

- Perfect starting guess: If $\boldsymbol{\Omega} = \mathbf{U}_1$ then

$$0 \le \log\det(\mathbf{I}+\mathbf{A}) - \log\det(\mathbf{I}+\mathbf{T}) = \underbrace{\log\det(\mathbf{I}+\boldsymbol{\Lambda}_2)}_{\textit{small eigenvalues}}$$

- General starting guess:

$$\begin{aligned} 0 \quad &\le \quad \log\det(\mathbf{I}+\mathbf{A}) - \log\det(\mathbf{I}+\mathbf{T}) \\ &\le \log\det\left(\mathbf{I}+\boldsymbol{\Lambda}_2\right) + \log\det\left(\mathbf{I}+\gamma^{2q-1}\,\|\boldsymbol{\Omega}_2\boldsymbol{\Omega}_1^{\dagger}\|^2\,\boldsymbol{\Lambda}_2\right) \end{aligned}$$

Absolute Error: Concentration

Starting guess $\mathbf{\Omega}$ is $n \times (k+p)$ standard Gaussian

For any $\delta > 0$, with probability at $1-\delta$

$$\begin{aligned} 0 \quad \leq \quad & \log\det(\mathbf{I}+\mathbf{A}) - \log\det(\mathbf{I}+\mathbf{T}) \\ & \leq \log\det\left(\mathbf{I}+\mathbf{\Lambda}_2\right) + \log\det\left(\mathbf{I} + c\,\gamma^{2q-1}\,\mathbf{\Lambda}_2\right) \end{aligned}$$

where

$$c \equiv \left(\frac{2}{\delta}\right)^{\frac{2}{p+1}} e^2 \, \frac{k+p}{(p+1)^2} \left(\sqrt{n-k} + \sqrt{k+p} + \sqrt{2\log\frac{2}{\delta}}\right)^2$$

Fast convergence with high probability:

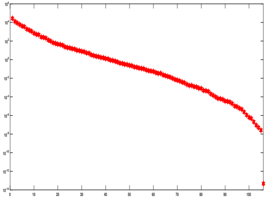
Failure probability $\delta \approx 10^{-15}$
Iterations $q \leq 2$, oversampling $p \approx 30$
Strong eigenvalue gap $\gamma \leq .8$

Numerical Experiments

Bayesian OED

$$\mathcal{H} \equiv C_{\text{prior}}^{1/2} \, F^* \, \Gamma_{\text{noise}}^{-1} \, F \, C_{\text{prior}}^{1/2}$$

Dimension $n = 1018$
rank$(\mathcal{H}) \approx 105$
Rapid eigenvalue decay:
$\lambda_1 \approx 10^4, \ \lambda_{105} \approx 10^{-8}$
$\lambda_{106} \approx 10^{-14}$
Iterations $q = 1$
Oversampling $p = 20$



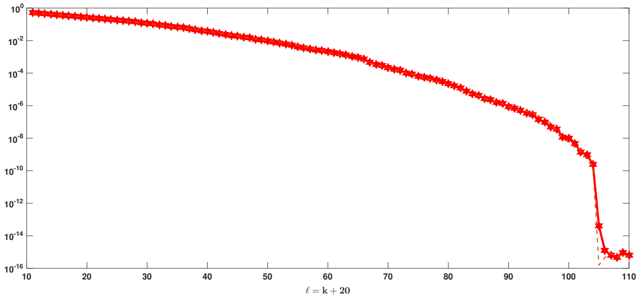
Estimators: $f(\circ) = \text{trace}(\circ)$ or $f(\circ) = \log\det(\mathbf{I} + \circ)$

Relative errors

$$\Delta \equiv \frac{f(\mathcal{H}) - f(\mathbf{T})}{f(\mathcal{H})}$$

Bayesian OED: trace($\mathcal{H}$) and $\log\det(I + \mathcal{H})$

Relative errors Δ vs subspace dimension k



Fast convergence, down to machine accuracy

Matrices of Small Dimension

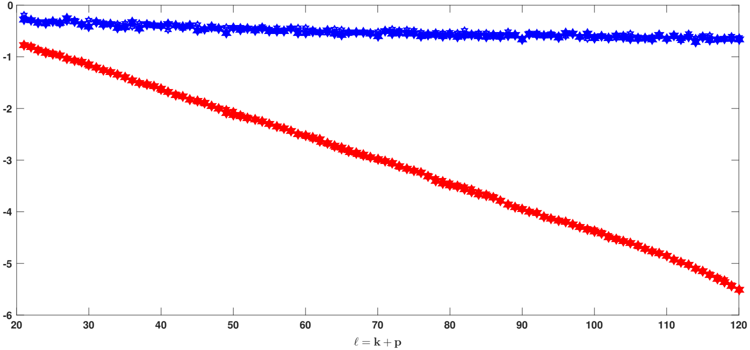
- Small matrix dimension: $n = 128$
- Geometrically decaying eigenvalues: $\lambda_j = \gamma^j \lambda_1 \qquad .86 \leq \gamma \leq .98$
- Oversampling: $p = 20$
- Estimators: $f(\circ) = \mathrm{trace}(\circ) \quad$ or $\quad f(\circ) = \log\det(\mathbf{I} + \circ)$

Relative errors

$$\Delta \equiv \frac{f(\mathbf{A}) - f(\mathbf{T})}{f(\mathbf{A})} \leq \left(1 + c\,\gamma^{2q-1}\right) \gamma^k \, \frac{1 - \gamma^{n-k}}{1 - \gamma^n}$$

Our Trace Estimator vs Monte Carlo

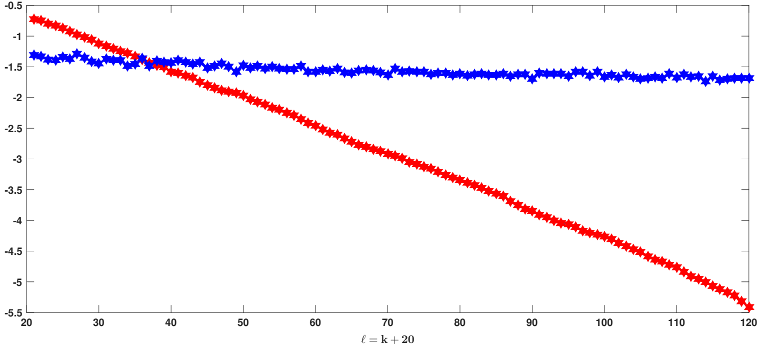
Relative errors $\log(\Delta)$ vs subspace dimension k



Our trace estimator much more accurate

Our Logdet Estimator vs Monte Carlo

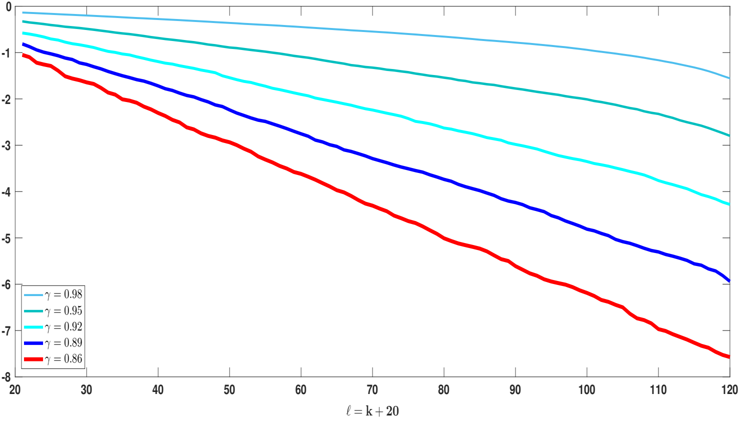
Relative errors $\log(\Delta)$ vs subspace dimension k



Our logdet estimator much more accurate

Effect of Eigenvalue Gap on Trace and Logdet

Relative errors $\log(\Delta)$ vs subspace dimension k



Our estimators more accurate as eigenvalue gap increases

Summary

Randomized trace and logdet estimators

for Hermitian positive semi-definite matrices

- Matrix-free (estimators use only matrix vector products)
- Random starting guesses: Gaussian and Rademacher
- Biased estimator
- Bayesian inverse problem: Fast convergence, high accuracy
- Much higher accuracy than Monte Carlo
- Error bounds informative even for small dimensions
- Clean analysis: first structural, then probabilistic